Solvation Layer of Antifreeze Proteins Analyzed with a Markov State Model

Wellig, Sebastian ; Hamm, Peter

Abstract: Three structurally very different antifreeze proteins (AFPs) are studied, addressing the question as to what extent the hypothesized preordering-binding mechanism is still relevant in the second solvation layer of the protein and beyond. Assuming a two-state model of water, the solvation layers are analyzed with the help of molecular dynamics simulations together with a Markov state model, which investigates the local tedrahedrality of the water hydrogen-bond network around a given water molecule. It has been shown previously that this analysis can discriminate the high-entropy, high-density state of the liquid (HDL) from its more structured low-density state (LDL). All investigated proteins, regardless of whether they are an AFP or not, have a tendency to increase the amount of HDL in their second solvation layer. The ice binding site (IBS) of the antifreeze proteins counteracts that trend, with either a hole in the HDL layer or a true excess of LDL. The results correlate to a certain extent with recent experiments, which have observed ice like vibrational (VSFG) spectra for the water atop the IBS of only a subset of antifreeze proteins. It is concluded that the preordering-binding mechanism indeed seems to play a role but is only part of the overall picture.

DOI: https://doi.org/10.1021/acs.jpcb.8b04491

Solvation Layer of Antifreeze Proteins Analyzed with a Markov State Model

Sebastian Wellig, Peter Hamm*
Department of Chemistry, University of Zurich, Switzerland
**corresponding author: peter.hamm@chem.uzh.ch*
(Dated: July 8, 2019)

Abstract: Three structurally very different antifreeze proteins (AFPs) are studied, addressing the question to what extent the hypothesized pre-ordering-binding mechanism is still relevant in the second solvation layer of the protein and beyond. Assuming a two-state model of water, the solvation layers are analyzed with the help of molecular dynamics simulations together with a Markov state model, which investigates the local tedrahedrality of the water hydrogen-bond network around a given water molecule. It has been shown previously that this analysis can discriminate the high-entropy, high-density state of the liquid (HDL) from its more structured low-density state (LDL). All investigated proteins, regardless whether they are an AFP or not, have a tendency to increase the amount of HDL in their second solvation layer. The ice binding site (IBS) of the antifreeze proteins counteracts that trend, with either a hole in the HDL layer or a true excess of LDL. The results correlate to a certain extent with recent experiments, which have observed ice-like vibrational (VSFG) spectra for the the water atop of the IBS of only a subset of anti-freeze proteins. It is concluded that the pre-ordering-binding mechanism indeed seems to play a role, but is only part of the overall picture.

I. INTRODUCTION

Antifreeze proteins (AFPs) are a class of proteins that protect organisms from ice formation at temperatures below freezing point.[1–4] They are found in polar fish, some insects, bacteria and plants, and are also of interest for the food industry, because they improve properties of frozen food with regard to texture or cryopreservation.[5,6] AFPs exhibit a rich structural variety with very different primary, secondary and tertiary structures,[2] and different evolutionary origins. The "adsorption-inhibition" mechanism is the accepted picture for AFPs. The active site of an anti-freeze proteins, the so-called ice-binding site (IBS), is a typically relative flat, rigid, and hydrophobic surface of the protein with often (but not always) a regular structure. The IBS binds to nascent ice crystals and thereby prevent their further growth.[4] AFPs introduce a thermal hysteresis gap, within which ice neither grows nor melts.

The mechanism, by which the binding to ice crystals exactly occurs, is however still debated. The originally suggested mechanism is purely based on H-bonding capabilities.[3] This hypothesis, however, fails to explain how AFPs are better binders than water, which is a capable H-bonder and in large excess (in the order of 1 mM of the AFP *versus* ≈55 M of water). Furthermore, since it was noticed that ice-binding sites are often rather hydrophobic, a new hypothesis based on the hydrophobic effect was suggested. That is, it is assumed that water on the ice-binding site forms clathrate-like structures, which have less degrees of freedom than bulk water. Ice-binding would then be entropically disfavoured to a lesser extent by the release of these clathrate-like solvation layers.[4,7] This idea is closely related to the "pre-ordering-binding" mechanism, according to which a locally tetrahedral solvation layer acts as a mediator between the IBS and an ice crystal.[7,8]

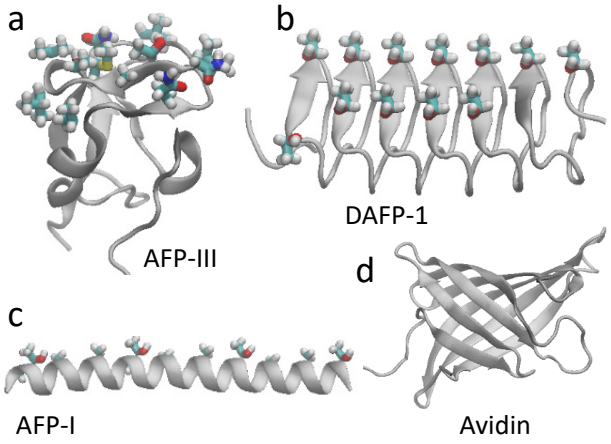

FIG. 1. Proteins considered in this study: (a) AFP-III (1MSI), (b) DAFP-1 (1EZG), (c) AFP-I (1WFA) and (d) avidin (3SZH). In the case of the antifreeze proteins, the side-chains of the IBS are indicated, i.e., for AFP-III: Q9, L10, I13, N14, T15, A16, T18, L19, V20, M21, V41, Q44,[9], for DAFP-1: T3, T16, T26, T28, T38, T40, T50, T52, T62, T64, T73, T76,[10] and for AFP-I: T2, A3, A6, A10, T13, A14, A17, A21, T24, A25, A28, A32, T35, A36.[11] Pictures were generated with VMD.[12]

A series of experimental papers in this regard has been published recently by Bakker and co-workers,[13–15] as well as by Weidner and co-workers,[16] who measured the vibrational sum frequency generation (VSFG) spectra of certain AFPs that accumulate and align at the water-air interface. Despite the fact that these experiments have been performed at room temperature, the VSFG band of the OH vibration of water is strongly red-shifted and resembles that of ice when type III AFP from an antarctic eelpout (abbreviated as AFP-III in the following, see Fig. 1a) is investigated.[13] The authors of this study therefore concluded the presence of locally tetra-

hedral water at the IBS of the protein. Interestingly, that spectroscopic feature completely disappeared in single point mutation of AFP-III (T18N),[13] which is known to be basically inactive as AFP.[17–19] On the other hand, they also did not observe the same ice-like VSFG spectrum for another AFP (DAFP-1, see Fig. 1b), despite the fact that it is more active as AFP than AFP-III, and concluded that the binding mechanism is different in this case.[14] Havenith and co-workers, on the other hand, have addressed the same question with the help of THz spectroscopy. They have seen subtle changes in the THz spectrum of water solvating DAFP-1, which they attributed to an extended solvation layer.[20] However, other studies from the same group suggested that this is a very universal observation for all proteins.[21,22]

Molecular dynamics (MD) simulations have been widely used to investigate the pre-ordering-binding mechanism. For instance, Sharp and co-workers suggested that the first water coordination shell around AFP-III indeed exhibits a slightly enhanced tetrahedrality at room temperature.[23,24] Similar conclusion were drawn by Daggett and co-workers[25], as well as Smith and coworkers,[8] using a variety of water force fields and order parameters to measure the amount of tetrahedrality. Approaching the problem from a different angle, Jana and coworkers used an free energy perspective, showing that the IBS of various AFBs "see" an ice surface over significant distances (up to 20 Å) in terms of a sloped potential of mean force.[26–28] On the other hand, Leitner and coworkers compared the partial specific heat of water near the IBS of DAFP-1 with the rest of the protein.[29] Both free energy and specific heat measure entropy from a thermodynmic rather than from a structural point of view.

The pre-ordering-binding mechanism, and in particular the observation of ice-like water atop of the IBS of AFP-III, suggests that there is a connection to the two-state model of water. That model has been put forward a long time ago[30] to explain the many anomalies in the thermodynamic properties of bulk water, and is closely related to the liquid-to-liquid phase transition hypothesis introduced by Stanley and coworkers[31,32]. According to this model, water can exist in two forms, a low-density liquid (LDL), which is locally tetrahedral up to the second coordination shell around each water molecule, *versus* a high-density liquid (HDL), which contains intercalated tetrahedral hydrogen bond networks.[32,33] It should be noted that alternative views on the structure of HDL have been proposed in literature, i.e., single-donor/single-acceptor chains or rings of water molecules.[34–36] In any case, while both forms of water are separated by a first-order phase transition below a liquid-to-liquid critical point (which is suspected at around 220 K), the transition becomes continuous above the liquid-to-liquid critical point with spatial domains of certain size that still can be characterized as LDL and HDL.[37,38] For example, the density maximum of water at 4°C is explained as a shifting equilibrium between LDL and HDL. Increasing computational[39–45] and experimental[46–54] evidence has recently been accumulated that suggests that such a liquid-to-liquid phase transition indeed exists for various computer models of water as well as for real water. From the computational point of view, the issue has been contested,[55–57] but much of that controversy has recently been resolved.[45]

It has been proposed that the dynamical transition of proteins (also sometimes called the "protein glass transition") is related to the liquid-liquid phase transition of the solvating water.[58] Here, we make a different connection between proteins and the two-state model of water. That is, the working hypothesis here is that the presence of a protein, and in particular the IBS of an AFP, locally shifts the equilibrium between HDL and LDL in the solvation layer of the protein. In fact, the pre-ordering-binding mechanism would suggest that the amount of LDL is enlarged at the IBS of an AFP.

In order to be able to identify HDL or LDL water, one needs a proper order-parameter. To the best of our knowledge, there exist only two options that work not only very close to the liquid-to-liquid critical point, but also at ambient temperatures. The first option is the so-called local structure index,[59–61] which considers all waters in the first and the second coordination shell around a central water molecule and evaluates their inter-molecular distances (for details see e.g. Eq. 1 of Ref. 60). The local structure index reveals a bimodal distribution corresponding to LDL and HDL, however only after quenching MD snapshots to 0 K.[60,61]

To avoid that drawback, we recently introduced a Markov state model as the second option for such an order parameter.[62] As for the local structure index, the first and second coordination shell around each water are analyzed, addressing the question whether it is locally tetrahedral up to the second coordination shell (resembling LDL) or whether the second coordination shell intercalates in the voids of the first one (HDL). To that end, the five closest waters around a given reference water are selected. While the four closest waters are tetrahedral for most of the time in any case, the position of the fifth-closest water, which is the first water in the second coordination shell, has been shown to be decisive in discriminating LDL from HDL.[63]

The basic idea of a Markov state model is to first discretize the space of local structures into highly resolved microstates by introducing a measure of structural similarity (such as an RMSD). One then determines the transition probability between each pair of microstates along a MD trajectory, and analyzes the "kinetic content" of that transition probability matrix.[64,65] The critical question is whether there exists a separation of timescales between the slowest and the subsequent faster timescales,[65] in which case the dynamics can be characterized as two-state. That is, fast timescales would indicate the dynamics within free energy basins, i.e. the pre-exponential factor k_0 in an Arrhenius law $k = k_0 e^{-F_a/k_B T}$ in the simplest possible scenario, while

the slowest component would reflect the rate k of hopping across the one dominant barrier separating two free energy basins. The Markov state model has higher resolution power than conventional local order parameters, such as the radial distribution function $g_5(r)$ of the fifth-closest water molecule,[63] allowing one to detect two-state behaviour even at room temperature in the presence of thermal noise, i.e., without having to quench MD simulation snapshots to 0 K. Using that approach, it has been shown in Ref. 62 that both bulk ST2[66] and TIP4P/2005[67] water indeed exhibit two-state behaviour in a wide temperature and pressure range from ambient conditions into the supercooled regime, and that the two corresponding subensembles can be characterized as LDL and HDL.

Here, we apply the same Markov state model to water solvating various AFPs. If the interpretation of Ref. 13 is correct with an ice-like (LDL) layer of water atop of the IBS of the protein, then this algorithm should detect it. For the purpose of this study, we selected AFP-III, its mutant T18N, DAFP-1 (Fig. 1a,b), as well as a type I AFP from winter flounder that consist of a single α-helix (abbreviated as AFP-I in the following, see Fig. 1c). With these examples, we cover most of the structural variability of AFPs.[2] As a counter example, we also investigated shwanavidin (abbreviated as avidin in the following), which is a β-barrel(Fig. 1d) similar to the IBS of DAFP-1. It does however not have any known activity as AFP.

II. METHODS

A. MD Simulations

PDB structures of the considered proteins (1MSI for AFP-III,[18] 1EZG for DAFP-1,[10] 1WFA for AFP-I,[68] and 3SZH for Avidin,[69] were converted into Gromacs[70] after (in some cases) extracting one protein from a unit cell with more than one protein, and after adding the amino acids that are missing in the X-ray structures, since they are in floppy regions (termini and loops) using MODELLER.[71] The mutant T18N of AFP-III has been constructed in the same way. The Gromacs implementation of the Charmm27 force field has been used.[72,73] The proteins were solvated with ≈7000 to ≈13000 water molecules, depending on the size of the protein, by defining a minimum distance of 10 Å from the protein to the faces of a cubic box (in the case of AFP-I, that layer was reduced to 7 Å in order to save computational time, since the molecule as a single α-helix is very large, albeit only in one dimension, and most of the time will not lie parallel to any of the box-edges). We used TIP4P/2005 water[67], despite the fact that the Charmm27 force field has been parametrized for TIP3P,[74] since TIP4P/2005 describes the thermodynamic properties of bulk water much better;[75] in fact we found that TIP3P does not reveal any two-state behavior based on the Markov state model. The simulation box was neutralized (if needed) with the proper number of Na^+ ions, energy minimized, and the water molecules pre-equilibrated for 1 ns with the protein constraint. For the subsequent molecular dynamics (MD) simulations, all bonds involving hydrogen atoms were constrained, allowing for a 2 fs time step. The long range electrostatic forces were treated by the Particle-Mesh-Ewald approximation[76] and the Lennard-Jones interactions were smoothly switched to zero from 9 Å to 10 Å. The simulations were performed in the NpT ensemble, with temperature (280 K, coupling time 0.2 ps) and pressure (1 bar, coupling time 0.5 ps) controlled using velocity-rescaling[77] and the Berendsen barostat,[78] respectively. Starting from the pre-equilibrated states, 1 μs long trajectories were launched with snapshots taken every 50 ns, where the first 50 ns served as equilibration. These snapshots were used as starting points for the actual production runs, each 1 ns long with a saving time of 20 fs. This interleaved protocol was chosen, since a simulation time of 1 ns was considered long enough to sample the water configurations around a given protein configuration, but too short to sample protein configurations. On the other hand, the full 1 μs trajectory would have been computationally too costly to be analyzed with the algorithm described in the next section.

B. Markov State Model

The Markov state model used to analyse of the protein solvation layer has been constructed in essentially the same way as in Ref. 62, where the two-state behaviour of bulk water has been investigated. In brief, the five closest water molecules around a given reference water molecule were selected. The continuous space of structures of these clusters of six water molecules was discretized into highly resolved microstates (typically 30'000-40'000 microstates), applying the hierarchical algorithm presented in Ref. 79 together with the RMSD[80] of the oxygen atoms used as a measure of structural similarity. Subsequently, a transition probability matrix between those microstates was calculated, whose eigenvalues represent the dynamical content of the system.

In comparison to Ref. 62, the algorithm had to be adapted only to a minor extent for the purpose of the present study. That is, since the structures of the five closest water molecules around a given water molecule need to be analyzed with regard to the question whether it is HDL or LDL-like, it does not make sense to consider reference water molecules in the first solvation layer around the protein, since those necessarily will have a smaller coordination number. We therefore defined a cut-off distance of 5 Å to the closest atom of the protein, which turned out to be the minimum between first and second solvation layer in a distance distribution function of waters around the protein (*vide infra*). We considered only water molecules beyond that cut-off in the construction of microstates and in the Markov state model. If the

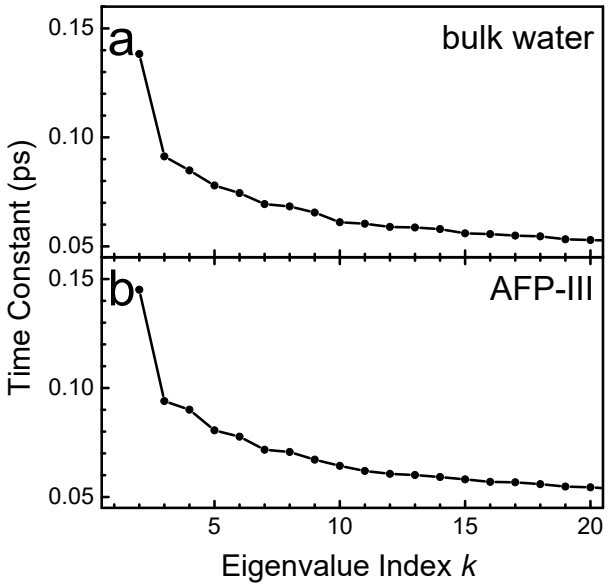

FIG. 2. The leading time constants of the Markov State Model for (a) bulk TIP4P/2005 water at 280 K and (b) TIP4P/2005 water solvating AFP-III. Panel (a) is shown for comparison and has been reproduced from Ref. 62.

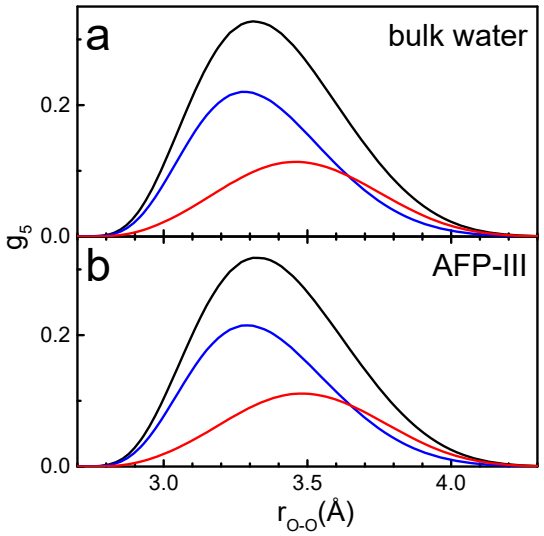

FIG. 3. Radial distribution functions $g_5(r)$ of the fifth-closest water molecule around a central reference water molecule for (a) bulk TIP4P/2005 water at 280 K and (b) TIP4P/2005 water solvating AFP-III. The black lines show the overall $g_5(r)$, while red and blue lines show the $g_5(r)$'s weighted with the two distributions $\rho_{LDL}^{(loc)}$ and $\rho_{HDL}^{(loc)}$. Panel (a) is shown for comparison and has been reproduced from Ref. 62.

simulation box contained ions (because the corresponding protein is charged), waters around the ions were discarded as well with the same cut-off.

From the eigenvector of the transition probability matrix corresponding to the equilibrium distribution (ρ_1, with time constant infinity) and the subsequent slowest kinetic process (ρ_2), an order-parameter:

$$p = \rho_2/\rho_1 \tag{1}$$

was calculated that discriminates LDL from HDL (see Figs. 1c and 3ef of Ref. 62). Furthermore, these eigenvectors were localized by an unitary transformation,[81] revealing two density distributions $\rho_{LDL}^{(loc)}$ and $\rho_{HDL}^{(loc)}$ (see Fig. 1d and Eqs. 7-11 of Ref. 62 for details), which were used as weight functions to calculate certain distribution functions of the LDL and HDL subensembles (they were assigned to LDL and HDL based on $g_5(r)$ discussed below). The "relative LDL excess" is defined as:

$$\epsilon_{LDL} \equiv \frac{\rho_{LDL}^{(loc)} - \rho_{HDL}^{(loc)}}{\rho_{LDL}^{(loc)} + \rho_{HDL}^{(loc)}}. \tag{2}$$

III. RESULTS

Figs. 2 and 3 compare the results of the Markov state model for bulk TIP4P/2005 water at 280 K (taken from Ref. 62) with those for TIP4P/2005 water solvating a protein, exemplified here for AFP-III (virtually the same results are obtained for all considered proteins). Fig. 2 shows the distribution of eigenvalues of the transition probability matrix, revealing a small but distinct separation of timescales between the slowest process (eigenvalue index k=2; the eigenvalue with index $k = 1$ is infinity and corresponds to the equilibrium distribution) and the subsequent faster processes (eigenvalue indices $k \geq 3$). This timescale separation evidences two-state behavior.

Fig. 3, in turn, shows the radial distribution function $g_5(r)$ of the fifth-closest water molecule of the overall simulation box (black) and of the two subensembles defined by the two localized distributions $\rho_{LDL}^{(loc)}$ (red) and $\rho_{HDL}^{(loc)}$ (blue). While the overall $g_5(r)$ is not bimodal and as such would not provide any criterion to discriminate LDL from HDL *per se* (it is bimodal only at temperatures very close to the liquid-to-liquid critical point),[63] the fact that the center-of-masses of the $g_5(r)$'s of the two sub-ensembles do differ still allows one to assign them to LDL or HDL. That is, the distribution shown in red peaks at 3.5 Å, indicating a fifth-closest water that is in a tetrahedral structure in the second coordination shell and hence the corresponding central reference water can be characterized as LDL (note that 3.5 Å is significantly closer than the 4.5 Å expected for the second coordination layer in a tetrahedral structure. That is since the fifth-closest water is the closest one from in total 12 waters in the second coordination layer. Due to that pre-selection and fluctuations in the liquid, its distribution peaks at a smaller distance). In contrast, the blue distribution peaks at 3.3 Å, indicating a fifth-closest water that is closer to the central water, since it intercalates in the voids formed by the tetrahedral structure of the first coordination shell, indicative for HDL. These $g_5(r)$'s together with 3D oxy-

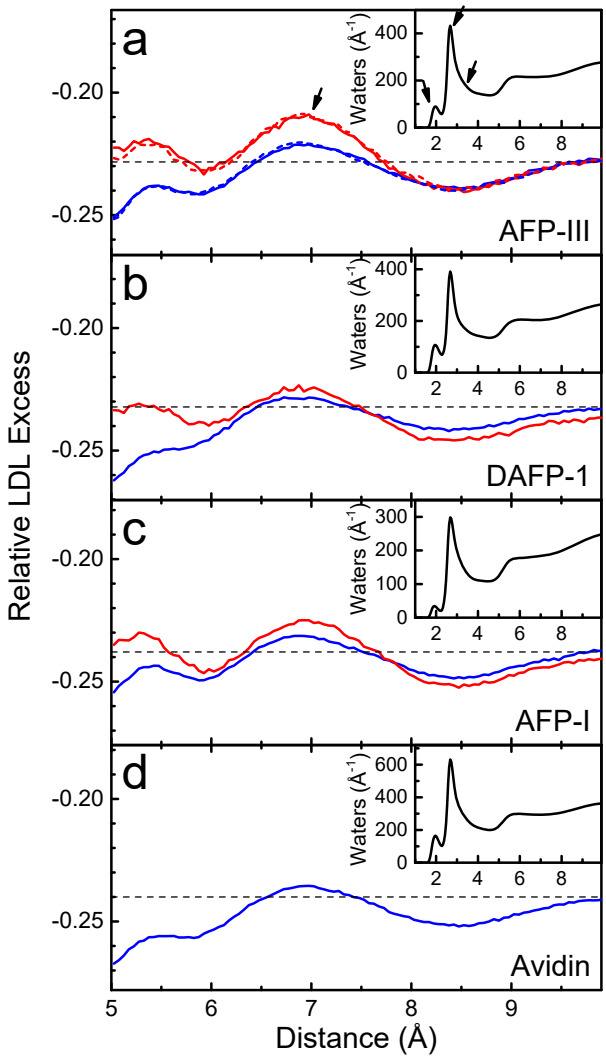

FIG. 4. Relative LDL excess ϵ_{LDL} as a function of distance to the protein for the three AFPs: (a) AFP-III (with the dotted line showing the result for the mutant T18N), (b) DAFP-1, (c) AFP-I and (d) avidin. Shown in blue is the ratio of all waters around the protein and in case of the AFPs in red the ratio of only the waters next to the active site (defined in the caption of Fig. 1). The dashed lines indicate ϵ_{LDL} averaged over the whole simulation box (which varies a little bit due to noise in the construction of the Markov state model). The inserts shows the overall distance distribution functions. The arrows mark features discussed in the text.

gen distributions (Fig. 4 in Ref. 62) have been used to verify that the Markov state model indeed discriminates LDL from HDL, even at ambient conditions.[62] The fact that the presence of the protein does not change these results evidences that the Markov state model is robust, or, put differently, that the simulation box is large enough to have sufficient number of water molecules with bulk properties that serve to calibrate the Markov state model.

The inserts of Figs. 4 show the number of waters as function of the distance from the protein for all considered proteins. The distance is defined as the distance of the oxygen atom of the central reference water molecule to the closest atom of the protein. These distance distribution functions are very similar for all considered proteins, regardless whether it is an AFP or not, with three features in the first solvation layer marked by the arrows in Fig. 4a (insert). The first and second peaks in the distance distribution functions at ~ 2 Å and ~ 3 Å correspond to waters that bind to the protein either as hydrogen-bond acceptors (with the closest atom of the proten being a proton) or as hydrogen-bond donors, respectively. We furthermore attribute the faint shoulder at ~ 4 Å to water molecules in the vicinity of hydrophobic parts of the proteins surface. The relative amplitudes of these features varies a little bit from protein to protein. The minimum at ~ 5 Å separates these waters from the second solvation layer and was therefore chosen as a cut-off for the Markov state model. The second solvation layers peaks at ~ 5.8 Å, while the distance distribution function is barely structured beyond.

The main panels of Fig. 4 show the relative LDL excess ϵ_{LDL} (see Eq. 2) as a function of distance, which is defined only beyond the cut-off of 5 Å. In bulk water at the temperature and pressure of the simulation, the amount of HDL is larger than that of LDL,[62] hence the "baseline" for ϵ_{LDL} is negative (see dashed lines in Fig. 4). What is relevant here is the extent to which ϵ_{LDL} deviates from that baseline (which still will be negative numbers). The blue lines in Fig. 4 show ϵ_{LDL} averaged over the whole protein surface. With regard to that, all proteins exhibit essentially the same trend, i.e., ϵ_{LDL} decreases as one approaches the protein surface, indicating an excess of HDL in the second solvation layer. For the AFPs, the additional red line in Fig. 4a-c show the relative LDL excess ϵ_{LDL} for only the waters in the vicinity of the IBS, which is defined by the amino acids shown in Fig. 1. The comparison between AFP-III and DAFP-1 is quite revealing. In both cases, the red line lies above the blue line in a range $\lesssim 8$ Å, indicating that the amount of LDL is larger in a relative sense. However, only in the case of AFP-III, the ϵ_{LDL} truly exceeds that of bulk water in a range from $\sim$5 Å to $\sim$7.5 Å (see arrow in the main panel of Fig. 4a; the dashed line indicates the value of ϵ_{LDL} averaged over the whole simulation box). In the case of DAFP-1, in contrast, ϵ_{LDL} oscillates a little bit, albeit only around the average value. AFP-I reveals a result that is somewhere in between AFP-III and DAFP-1. The T18N mutant of AFP-III is essentially indistinguishable from AFP-III (see Fig. 4a, dotted lines), despite the fact that this single point mutation essentially switches off its activity as AFP,[17–19] and it differs significantly from AFP-III with regard to its VSFG spectrum.[13]

The significance of the difference between AFP-III and DAFP-1 becomes clearer in Fig. 5a,b, which shows the 3D distributions of ϵ_{LDL} around the protein. These plots where generated by aligning all MD snapshots onto the average backbone structure by minimizing the RMSD of the C_α atoms. As the protein was aligned, all water molecules were translated and rotated in the same way.

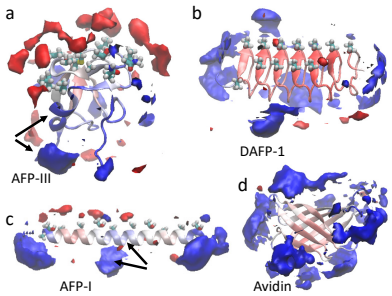

FIG. 5. Distribution of the relative LDL excess ϵ_{LDL} around (a) AFP-III, (b) DAFP-1, (c) AFP-I, and (d) avidin. The red contour surfaces are at $\epsilon_{LDL} = -0.205$, indicating an excess of LDL relative to the bulk value, and the blue contour surfaces at $\epsilon_{LDL} = -0.27$ indicating an excess of HDL. The residue of the IBS (in the case of the AFPs, panels a-c), and the protein backbone are shown as well. The latter is colored according to rigidity: red for more rigid over white towards blue for more flexible. The arrows mark features discussed in the text. Pictures were generated with VMD.[12]

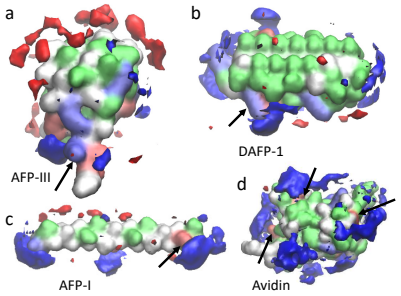

FIG. 6. Same as Fig. 5, but the protein now shown as surface plot (oriented in the same way as in Fig. 5) and colored according to residuum-type: Non-polar residues in white, polar residues in green, basic residues in blue, and acidic residues red. The arrows mark charged amino acids. Pictures were generated with VMD.[12]

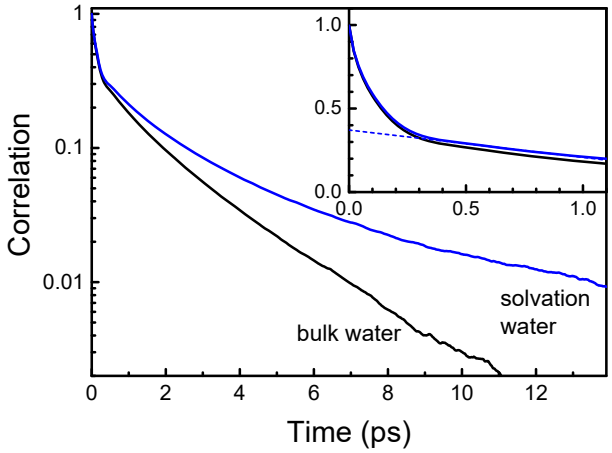

FIG. 7. Time-correlation function for bulk water (black) and the water molecules around AFP-III that remain in a solvation layer between 5 Å and 8 Å during the calculation of the correlation function (blue). The insert highlights the initial drop of the correlation function, with the dashed line extrapolating the slow component to $t = 0$. The bulk water data have reproduced from Ref. 62.

Densities were calculated by binning into cubes of size 1 Å^3 and by subsequent Gaussian smoothing with width 1.5 Å. Only bins containing $> 3\%$ of the bulk water density are shown in order to suppress the noise that occurs close to the protein when the water content becomes low. While we see an excess of LDL atop of the IBS in the case of AFP-III (Fig. 5a), indicated by the red contour surface, we find only a "hole" atop of the IBS of DAFP-1 in an otherwise dominating HDL layer shown in blue (Fig. 5b). DAFP-1 and avidin are not qualitatively different with regard to the 3D distributions of ϵ_{LDL}, despite the fact that avidin has no known activity as AFP. That is, also avidin exhibits HDL patches with similar-sized holes on the protein surface (Fig. 5d). Finally, just like for Fig. 4a-c, AFP-I lies in between AFP-III and DAFP-1 with some excess of LDL (shown in red) atop of the IBS.

Also depicted in Fig. 5 is the flexibility of protein backbone encoded by its color (red for more rigid, over white towards blue for more flexible). There is a clear correlation between flexibility of the backbone and the amount of HDL in the solvation layer. This is most evident for AFP-I (Fig. 5c), which as a single α-helix is quite flexible towards its termini with two pronounced patches of HDL around them. But there is a region of higher flexibility also in the middle of the helix accompanied by another patch of HDL, both of which are marked by an arrow in Fig. 5c. This correlation is harder to see in the other AFPs but does exist as well, see e.g. arrows in Fig. 5a for AFP-III. Conversely, the IBS of DAFP-1 (Fig. 5b) is the most rigid structure among all proteins studied here, and consequently no excess of HDL is formed.

Fig. 6 shows the same as Fig. 5, but with protein now colored according to residuum-type. Holes in the HDL layer coincide with regions with apolar (white) or at most polar (green) amino acids at the surface, while charged amino acids tend to induce HDL patches (see arrows). That observation is very universal and is true for all AFPs as well as for the control avidin.

We furthermore consider the time-correlation function:

$$c(t) \equiv \langle p(0)p(t) \rangle, \tag{3}$$

where the order parameter $p(t)$ is defined in Eq. 1. Fig.7, black line, shows this correlation function for bulk water,[62] while the blue line shows it for only those water molecules around AFP-III that remain in a solvation layer between 5 Å and 8 Å in the time-window of the correlation function. In either case, the time-correlation function decays in a biphasic manner, with an initial very quick drop and as subsequent much slower, close-to-exponential tail. It has been shown in Ref. 62 that this exponential tail reflects the diffusive motion of water molecules in and out from spatial domains, in which either LDL or HDL dominates. When going from bulk

water to the solvation layer of AFP-III, that diffusive process becomes somewhat slower (Fig.7, blue *versus* black line), since the position of these domains are now fixed to a certain extent by the protein (Fig. 5a). On the one hand, that resembles the fact that the diffusivity of water in the solvation layer of a protein is reduced, but in addition to that, it is now only the motion of the water molecules, no longer that of the domains, which determines the long-time decay of the correlation function.

The ultrafast initial drop of the time-correlation function reflects thermal noise in the analysis of the Markov state model. This in turn implies that the relative LDL excess ϵ_{LDL} shown in Figs. 4 and 5, which averages over time and hence over that ultrafast component, cannot reveal the limiting values ± 1 that Eq. 2 would suggest, even in the very best case. If we assume that the biphasic decay of the correlation function originates from a fast Gaussian process (thermal noise) and a slow binary process (switching between HDL and LDL), both of which are uncorrelated, then the maximal possible amplitude would be $\pm\sqrt{0.37} = \pm 0.6$ (where 0.37 is the amplitude of the diffusive process, see dashed line in Fig. 7, insert). On the other hand, the observed peak amplitudes in the data of Fig. 5a are $\Delta\epsilon_{LDL} \approx \pm 6\%$, (the peak amplitudes are smaller in Fig. 4a since these data are averaged over the whole IBS surface). If one normalizes out the contribution of thermal noise, the domains depicted in color in Fig. 5a contain an excess of about $\approx \pm 10\%$ LDL or HDL.

IV. DISCUSSION

All proteins investigated here, regardless whether they are an AFP or not, locally change the equilibrium between the two states of water around them, and have a tendency to increase the amount of HDL in the second solvation layer relative to that in the bulk. This observation seems to reflect the notion that solvation water around a protein is more dense than bulk water.[82–85] In essence, the presence of a protein surface disturbs the hydrogen bond network of water and therefore makes its higher-entropy form –HDL– more likely. The fact that the amount of HDL correlates with the flexibility of the protein backbone supports this view.

The IBS of AFPs counteracts that trend, a result which has been observed for all AFPs studied here that cover most of the structural variability of AFPs.[2] While the amount LDL does not necessarily exceed that of bulk water, it is still always larger than the LDL content averaged over the complete protein surface. That is, we observe either a "hole" in the HDL layer atop of the IBS (DAFP-1, Fig. 5b), or a true excess of the more structured LDL relative to bulk water (AFP-III, Fig. 5a and AFP-I, Fig. 5c), with the effect being more pronounced for AFP-III. With 7 apolar amino acids out of 12 (Fig. 6a), the IBS of AFP-III is relatively hydrophobic, suggesting that it is the hydrophobicity which enhances the preference of LDL, along the lines of the proposed clathrate-like structures around the IBS.[4,7] In contrast, the side-chains of the IBS of DAFP-1 are all hydrogen-bonding (polar) threonines (Fig. 6b).

Our results correlate to a certain extent with the observations of Bakker and coworkers, who observed ice-like water in VSFG spectra only in the case of wild-type AFP-III,[13] but not for DAFP-1,[14] despite the fact that the latter has the larger activity as AFP. It should however be stressed that the excess of LDL we do observe ($\approx +10\%$) is too small to call it ice-like, and we certainly would not expect that the IR spectrum of this water resembles that of ice. Furthermore, we observe no effect for the single-point mutation T18N of AFP-III, which is known to be basically inactive as AFP,[17–19] and for which the VSFG spectroscopy changed dramatically as well.[13] In comparing our results with those of Bakker and coworkers, one should however keep in mind that we had to exclude the first solvation layer from the Markov State analysis, while presumably the contribution of exactly that layer dominates the VSFG spectrum. However, structure in water has a correlation length,[62] and some of the structuring in the first solvation layer will extend into the second, which would be seen in the VSFG as well. If the interpretation of "ice-like water" is correct in Ref.[13], it is hard to imagine unstructured water in the second layer, while the first is "ice-like". Furthermore, they see an ice-like spectrum only at higher AFP-III concentrations with a relatively sharp transition as a function of concentration.[13] This observation implies that it is a cooperative effect in some way, which is not accounted for in the present study that contains only a single protein molecule. It might well be that this cooperative effect would occur only at a higher concentration in the T18N mutant of AFP-III. In addition, the VSFG spectroscopy selectively measures proteins that accumulate at the water-air interface, the latter of which is known to change the properties of water, again in a way that is not accounted for here. Finally, ice formation and ice growth might be a very slow process, and we miss it in our 1 μs long MD simulation. In that regard, it should however be mentioned that within signal-to-noise, we could not detect any time-evolution of the effect shown in Fig. 4a. Provided that the IBS could be considered a nucleation site similar to an ice crystal, ice growth has been shown to proceed on a 10-100 ns timescale,[86] i.e., within our time-window. As an alternative explanation, it has been argued that TIP4P/2005 (and all other computer models of water) underestimates local fluctuations of water, and reproduce them, if at all, only at significantly lower temperatures.[87,88] This correlates with the observation that also the freezing point of TIP4P/2005 is severely underestimated.[86] All these points might explain why we see the same trend as Bakker and coworker in their VSFG spectra,[13,14] albeit much less distinct.

Concerning the distribution of HDL around the protein, there is no qualitative difference between DAFP-1 and avidin (Fig. 5b versus Fig. 5d), yet avidin has no

known activity as AFP. Hence, the pre-ordering-binding mechanism, which we do indeed observe, cannot be the complete story. Other aspects are the planarity and the rigidity of the IBS, which are different for DAFP-1 and avidin. Another example in this regard is the dramatic effect of the single-point mutation T18N of AFP-III, reducing the activity to 10% of that of the wild-type, an effect which has been attributed to the increased side-chain length and side-chain volume that very locally leads to a steric disturbance of the IBS.[17,18] In terms of polarity, however, threonine and asparagine are very much comparable, explaining why there is no effect on the water structure in the second solvation layer and beyond (Fig. 4a).

V. CONCLUSION

In conclusion, we propose a direct link between the hypothesized pre-ordering-binding mechanism of AFPs and the two-state model of water. While the construct of the order parameter implied that we had to restrict the analysis to the second solvation layer, in contrast to most previous simulation studies that concentrated on the first solvation layer,[8,23–25] the structural resolution power of the Markov state model is large enough to see the quite subtle changes in the LDL/HDL distribution in the second solvation layer and beyond. The analysis reveals that all AFPs exhibit an excess of LDL atop of the IBS compared to the rest of the protein, while polar and/or more flexible amino acids tend to increase the HDL content. On the one hand, the results give insight into the binding mechanism of AFPs, in the sense that the pre-ordering-binding mechanism indeed seems to play a role, but is only part of the overall picture. At the same time, the results also serve to better understand the properties of water; in particular with regard to the question how they are changed in the vicinity of a protein. It is the interplay between water and protein that determines the properties of the latter to a large extent, a problem that has been studied since decades[89] and still is a matter of intense research.[90]

Acknowledgments: The work has been supported by the Swiss National Science Foundation (SNF) through the NCCR MUST as well as Grant 200021_165789/1.

References:

(1) Duman, J. G. Antifeeze and icenucteator proteins in tessertrial arthropoids, *Annu. Rev. Physiol.* **2001**, *63*, 327–57.
(2) Jia, Z.; Davies, P. L. Antifreeze proteins: An unusual receptor-ligand interaction, *Trends Biochem. Sci.* **2002**, *27*, 101–106.
(3) Davies, P. L.; Baardsnes, J.; Kuiper, M. J.; Walker, V. K. Structure and function of antifreeze proteins, *Philos. Trans. R. Soc. B Biol. Sci.* **2002**, *357*, 927–935.
(4) Davies, P. L. Ice-binding proteins: A remarkable diversity of structures for stopping and starting ice growth, *Trends Biochem. Sci.* **2014**, *39*, 548–555.
(5) Clarke, C. J.; Buckley, S. L.; Lindner, N. Ice structuring proteins-a new name for antifreeze proteins, *Cryoletters* **2002**, *23*, 89–92.
(6) Halwani, D. O.; Brockbank, K. G.; Duman, J. G.; Campbell, L. H. Recombinant Dendroides canadensis antifreeze proteins as potential ingredients in cryopreservation solutions, *Cryobiology* **2014**, *68*, 411–418.
(7) Garnham, C. P.; Campbell, R. L.; Davies, P. L. Anchored clathrate waters bind antifreeze proteins to ice, *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 7363–7367.
(8) Nutt, D. R.; Smith, J. C. Dual function of the hydration layer around an antifreeze protein revealed by atomistic molecular dynamics simulations, *J. Am. Chem. Soc.* **2008**, *130*, 13066–13073.
(9) Garnham, C. P.; Natarajan, A.; Middleton, A. J.; Kuiper, M. J.; Braslavsky, I.; Davies, P. L. Compound ice-binding site of an antifreeze protein revealed by mutagenesis and fluorescent tagging, *Biochemistry* **2010**, *49*, 9063–9071.
(10) Liou, Y. C.; Tocilj, A.; Davies, P. L.; Jia, Z. Mimicry of ice structure by surface hydroxyls and water of a beta-helix antifreeze protein., *Nature* **2000**, *406*, 322–324.
(11) Baardsnes, J.; Jelokhani-Niaraki, M.; Kondejewski, L. H.; Kuiper, M. J.; Kay, C. M.; Hodges, R. S.; Davies, P. L. Antifreeze protein from shorthorn sculpin: identification of the ice-binding surface., *Protein Sci.* **2001**, *10*, 2566–2576.
(12) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics, *J. Mol. Graph.* **1996**, *14*, 33–38.
(13) Meister, K.; Strazdaite, S.; DeVries, A. L.; Lotze, S.; Olijve, L. L. C.; Voets, I. K.; Bakker, H. J. Observation of ice-like water layers at an aqueous protein surface, *Proc. Natl. Acad. Sci.* **2014**, *111*, 17732–17736.
(14) Meister, K.; Lotze, S.; Olijve, L. L. C.; Devries, A. L.; Duman, J. G.; Voets, I. K.; Bakker, H. J. Investigation of the ice-binding site of an insect antifreeze protein using sum-frequency generation spectroscopy, *J. Phys. Chem. Lett.* **2015**, *6*, 1162–1167.
(15) Groot, C. C.; Meister, K.; Devries, A. L.; Bakker, H. J. Dynamics of the hydration water of antifreeze glycoproteins, *J. Phys. Chem. Lett.* **2016**, *7*, 4836–4840.
(16) Pandey, R.; Usui, K.; Livingstone, R. A.; Fischer, S. A.; Pfaendtner, J.; Backus, E. H.; Nagata, Y.; Fröhlich-Nowoisky, J.; Schmüser, L.; Mauri, S.; Scheel, J. F.; Knopf, D. A.; Pöschl, U.; Bonn, M.; Weidner, T. Ice-nucleating bacteria control the order and dynamics of interfacial water, *Sci. Adv.* **2016**, *2*, e1501630.
(17) Chao, H.; DeLuca, C. I.; Davies, P. L.; Sykes, B. D.; Sönnichsen, F. D. Structure-function relationship in the globular type III antifreeze protein: Identification of a

cluster of surface residues required for binding to ice, *Protein Sci.* **1994**, *3*, 1760–1769.
(18) Jia, Z.; DeLuca, C. I.; Chao, H.; Davies, P. L. Structural basis for the binding of a globular antifreeze protein to ice, *Nature* **1996**, *384*, 285–288.
(19) Graether, S.; DeLuca, C.; Baardsnes, J. Quantitative and qualitative analysis of type III antifreeze protein structure and function, *J. Biol. Chem.* **1999**, *274*, 11842–11847.
(20) Meister, K.; Ebbinghaus, S.; Xu, Y.; Duman, J. G.; DeVries, A.; Gruebele, M.; Leitner, D. M.; Havenith, M. Long-range protein-water dynamics in hyperactive insect antifreeze proteins, *Proc. Natl. Acad. Sci USA* **2013**, *110*, 1617–1622.
(21) Ebbinghaus, S.; Kim, S. J.; Heyden, M.; Yu, X.; Heugen, U.; Gruebele, M.; Leitner, D. M.; Havenith, M. An extended dynamical hydration shell around proteins, *Proc. Natl. Acad. Sci USA* **2007**, *104*, 20749–20752.
(22) Xu, Y.; Havenith, M. Perspective: Watching low-frequency vibrations of water in biomolecular recognition by THz spectroscopy, *J. Chem. Phys.* **2015**, *143*, 170901.
(23) Gallagher, K. R.; Sharp, K. A. Analysis of thermal hysteresis protein hydration using the random network model, *Biophys. Chem.* **2003**, *105*, 195–209.
(24) Yang, C.; Sharp, K. A. The mechanism of the type III antifreeze protein action: A computational study, *Biophys. Chem.* **2004**, *109*, 137–148.
(25) Smolin, N.; Daggett, V. Formation of ice-like water structure on the surface of an antifreeze protein, *J. Phys. Chem. B* **2008**, *112*, 6193–6202.
(26) Chakraborty, S.; Jana, B. Molecular insight into the adsorption of spruce budworm antifreeze protein to an ice surface: A clathrate-mediated recognition mechanism, *Langmuir* **2017**, *33*, 7202–7214.
(27) Chakraborty, S.; Jana, B. Conformational and hydration properties modulate ice recognition by type I antifreeze protein and its mutants, *Phys. Chem. Chem. Phys.* **2017**, *19*, 11678–11689.
(28) Chakraborty, S.; Jana, B. Optimum number of anchored clathrate water and its instantaneous fluctuations dictate ice plane recognition specificities of insect antifreeze protein, *J. Phys. Chem. B* **2018**, *122*, 3056–3067.
(29) Pandey, H. D.; Leitner, D. M. Thermodynamics of hydration water around an antifreeze protein: A molecular simulation study, *J. Phys. Chem. B* **2017**, *121*, 9498–9507.
(30) Röntgen, W. C. Ueber die Constitution des flüsigen Wassers, *Ann. der Phys. und Chemie, N.F.* **1892**, *45*, 91–97.
(31) Poole, P.; Sciortino, F.; Essmann, U.; Stanley, H. Phase behaviour of metastable water, *Nature* **1992**, *360*, 324.
(32) Mishima, O.; Stanley, H. E. The relationship between liquid, supercooled and glassy water, *Nature* **1998**, *396*, 329–335.
(33) Soper, A. K. Structures of High-Density and Low-Density Water, *Phys. Rev. Lett.* **2000**, *84*, 2881–2884.
(34) Wernet, P.; Nordlund, D.; Bergmann, U.; Cavalleri, M.; Odelius, M.; Ogasawara, H.; Näslund, L. Å. A.; Hirsch, T. K.; Ojamäe, L.; Glatzel, P.; Pettersson, L. G. M.; Nilsson, A. The structure of the first coordination shell in liquid water, *Science* **2004**, *304*, 995—-999.
(35) Harada, Y.; Tokushima, T.; Horikawa, Y.; Takahashi, O.; Niwa, H.; Kobayashi, M.; Oshima, M.; Senba, Y.; Ohashi, H.; Wikfeldt, K. T.; Nilsson, A.; Pettersson, L. G.; Shin, S. Selective probing of the OH or OD stretch vibration in liquid water using resonant inelastic soft-X-ray scattering, *Phys. Rev. Lett.* **2013**, *111*, 193001.
(36) Harada, Y.; Miyawaki, J.; Niwa, H.; Yamazoe, K.; Pettersson, L. G.; Nilsson, A. Probing the OH stretch in different local environments in liquid water, *J. Phys. Chem. Lett.* **2017**, *8*, 5487–5491.
(37) Nilsson, A.; Pettersson, L. G. M. Perspective on the structure of liquid water, *Chem. Phys.* **2011**, *389*, 1–34.
(38) Gallo, P.; Amann-Winkel, K.; Angell, C. A.; Anisimov, M. A.; Caupin, F.; Chakravarty, C.; Lascaris, E.; Loerting, T.; Panagiotopoulos, A. Z.; Russo, J.; et al. Water: A tale of two liquids, *Chem. Rev.* **2016**, *116*, 7463–7500.
(39) Liu, Y.; Palmer, J. C.; Panagiotopoulos, A. Z.; Debenedetti, P. G. Liqui-liquid transition in ST2 water, *J. Chem. Phys.* **2012**, *137*, 214505.
(40) Poole, P. H.; Bowles, R. K.; Saika-Voivod, I.; Sciortino, F. Free energy surface of ST2 water near the liquid-liquid phase transition, *J. Chem. Phys.* **2013**, *138*, 34505.
(41) Palmer, J. C.; Martelli, F.; Liu, Y.; Car, R.; Panagiotopoulos, A. Z.; Debenedetti, P. G. Metastable liquid-liquid transition in a molecular model of water, *Nature* **2014**, *510*, 385–388.
(42) Smallenburg, F.; Filion, L.; Sciortino, F. Erasing no-man's land by thermodynamically stabilizing the liquid-liquid transition in tetrahedral particles, *Nat. Phys.* **2014**, *10*, 653–657.
(43) Singh, R. S.; Biddle, J. W.; Debenedetti, P. G.; Anisimov, M. A. Two-state thermodynamics and the possibility of a liquid-liquid phase transition in supercooled TIP4P/2005 water, *J. Chem. Phys.* **2016**, *144*, 144504.
(44) Ni, Y.; Skinner, J. L. IR spectra of water droplets in no man's land and the location of the liquid-liquid critical point, *J. Chem. Phys.* **2016**, *145*, 124509.
(45) Palmer, J. C.; Haji-Akbari, A.; Singh, R.; Martelli, F.; Car, R.; Panagiotopoulos, A. Z.; Debenedetti, P. G. Comment on "The putative liquid-liquid transition is a liquid-solid transition in atomistic models of water" [I and II: J. Chem. Phys. 135, 134503 (2011); J. Chem. Phys. 138, 214504 (2013)], *J. Chem. Phys.* **2018**, *148*, 137101.
(46) Scherer, J. R.; Go, M. K.; Kint, S. Raman spectra and structure of water from -10 to 90 degC, *J. Phys. Chem.* **1974**, *78*, 1304–1313.
(47) Huang, C.; Wikfeldt, K. T.; Tokushima, T.; Nordlund, D.; Harada, Y.; Bergmann, U.; Niebuhr, M.; Weiss, T. M.; Horikawa, Y.; Leetmaa, M.; et al. The inhomogeneous structure of water at ambient conditions, *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 15214–15218.
(48) Maréchal, Y. The molecular structure of liquid water delivered by absorption spectroscopy in the whole IR region completed with thermodynamics data, *J. Mol. Struct.* **2011**, *1004*, 146–155.
(49) Manka, A.; Pathak, H.; Tanimura, S.; Wölk, J.; Strey, R.; Wyslouzil, B. E. Freezing water in no-man's land, *Phys. Chem. Chem. Phys.* **2012**, *14*, 4505.
(50) Sellberg, J. A.; Huang, C.; McQueen, T. A.; Loh, N. D.; Laksmono, H.; Schlesinger, D.; Sierra, R. G.; Nordlund, D.; Hampton, C. Y.; Starodub, D.; et al. Ultrafast X-ray probing of water structure below the homogeneous ice nucleation temperature, *Nature* **2014**, *510*, 381–384.
(51) Amann-Winkel, K.; Gainaru, C.; Handle, P. H.; Seidl, M.; Nelson, H.; Böhmer, R.; Loerting, T. Water's second

glass transition, *Proc. Natl. Acad. Sci. USA* **2014**, *110*, 17720–17725.
(52) Perakis, F.; Amann-Winkel, K.; Lehmkühler, F.; Sprung, M.; Mariedahl, D.; Sellberg, J. A.; Pathak, H.; Späh, A.; Cavalca, F.; Schlesinger, D.; et al. Diffusive dynamics during the high-to-low density transition in amorphous ice, *Proc. Natl. Acad. Sci.* **2017**, *114*, 8193–8198.
(53) Kim, K. H.; Späh, A.; Pathak, H.; Perakis, F.; Mariedahl, D.; Amann-Winkel, K.; Sellberg, J. A.; Lee, J. H.; Kim, S.; Park, J.; et al. Maxima in the thermodynamic response and correlation functions of deeply supercooled water, *Science* **2017**, *358*, 1589–1593.
(54) Morawietz, T.; Marsalek, O.; Pattenaude, S. R.; Streacker, L. M.; Ben-Amotz, D.; Markland, T. E. The Interplay of Structure and Dynamics in the Raman Spectrum of Liquid Water over the Full Frequency and Temperature Range, *J. Phys. Chem. Lett.* **2018**, *9*, 851.
(55) Limmer, D. T.; Chandler, D. The putative liquid-liquid transition is a liquid-solid transition in atomistic models of water., *J. Chem. Phys.* **2011**, *135*, 134503.
(56) Limmer, D.; Chandler, D. The putative liqui-liquid phase transition is a liquid-solid transition in atomistic models of water, part II, *J. Chem. Phys.* **2013**, *138*, 214504.
(57) Chandler, D. Metastability and no criticality, *Nature* **2016**, *513*, E1—-E2.
(58) Kumar, P.; Yan, Z.; Xu, L.; Mazza, M. G.; Buldyrev, S. V.; Chen, S. H.; Sastry, S.; Stanley, H. E. Glass transition in biomolecules and the liquid-liquid critical point of water, *Phys. Rev. Lett.* **2006**, *97*, 177802.
(59) Shiratani, E.; Sasai, M. Growth and collapse of structural patterns in the hydrogen bond network in liquid water, *J. Chem. Phys.* **1996**, *104*, 7671–7680.
(60) Appignanesi, G. A.; Fris, J. A. R.; Sciortino, F. Evidence of a two-state picture for supercooled water and its connections with glassy dynamics, *Eur. Phys. J. E* **2009**, *29*, 305–310.
(61) Wikfeldt, K. T.; Nilsson, A.; Pettersson, L. G. M. Spatially inhomogeneous bimodal inherent strcuture of simulated water, *Phys. Chem. Chem. Phys.* **2011**, *13*, 19918–19924.
(62) Hamm, P. Markov state model of the two-state behaviour of water, *J. Chem. Phys.* **2016**, *145*, 134501.
(63) Cuthbertson, M. J.; Poole, P. H. Mixturelike behavior near a liquid-liquid phase transition in simulations of supercooled water, *Phys. Rev. Lett.* **2011**, *106*, 1–4.
(64) Pande, V. S.; Beauchamp, K.; Bowman, G. R. Everything you wanted to know about Markov state models but were afraid to ask., *Methods* **2010**, *52*, 99–105.
(65) Prinz, J.-H.; Wu, H.; Sarich, M.; Keller, B.; Senne, M.; Held, M.; Chodera, J. D.; Schütte, C.; Noé, F. Markov models of molecular kinetics: Generation and validation, *J. Chem. Phys.* **2011**, *134*, 174105.
(66) Stillinger, F. H.; Rahman, A. Improved simulation of liquid water by molecular dynamics, *J. Chem. Phys.* **1974**, *60*, 1545–1957.
(67) Abascal, J. L.; Vega, C. A general purpose model for the condensed phases of water: TIP4P/2005., *J. Chem. Phys.* **2005**, *123*, 234505.
(68) Ice-binding structure and mechanism of an antifreeze protein from winter flounder. Sicheri, F.; Yang, D. S. C. **1995**.
(69) Meir, A.; Bayer, E. A.; Livnah, O. Structural adaptation of a thermostable biotin-binding protein in a psychrophilic environment, *J. Biol. Chem.* **2012**, *287*, 17951–17962.
(70) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. Gromacs; fast, flexible and free, *J. Comput. Chem.* **2005**, *26*, 1701–1718.
(71) Webb, B.; Sali, A. Comparative protein structure modeling using MODELLER, *Curr. Protoc. Bioinforma.* **2014**, *47*, 5.6.1–5.6.32.
(72) MacKerell, A. D.; Feig, M.; Brooks III, C. L. Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations, *J. Comput. Chem.* **2004**, *25*, 1400–1415.
(73) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins, *J. Phys. Chem. B.* **1998**, *102*, 3586–3616.
(74) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water, *J. Chem. Phys.* **1983**, *79*, 926–935.
(75) Vega, C.; Abascal, J. L. F. Simulating water with rigid non-polarizable models: a general perspective, *Phys. Chem. Chem. Phys.* **2011**, *13*, 19663.
(76) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An Nlog(N) method for Ewald sums in large systems, *J. Chem. Phys.* **1993**, *98*, 10089–10092.
(77) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling, *J. Chem. Phys.* **2007**, *126*, 014101.
(78) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath, *J Chem Phys* **1984**, *81*, 3684–3690.
(79) Vitalis, A.; Caflisch, A. Efficient construction of mesostate networks from molecular dynamics trajectories, *J. Chem. Theory Comput.* **2012**, *8*, 1108–1120.
(80) Coutsias, E. A.; Seok, C.; Dill, K. A. Using quaternions to calculate RMSD, *J. Comput. Chem.* **2004**, *25*, 1849–1857.
(81) Edmiston, C.; Ruedenberg, K. Localized atomic and molecular orbitals, *Rev. Mod. Phys.* **1963**, *35*, 457–465.
(82) Svergun, D. I.; Richard, S.; Koch, M. H.; Sayers, Z.; Kuprin, S.; Zaccai, G. Protein hydration in solution: experimental observation by x-ray and neutron scattering., *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 2267–2272.
(83) Merzel, F.; Smith, J. C. Is the first hydration shell of lysozyme of higher density than bulk water?, *Proc. Natl. Acad. Sci.* **2002**, *99*, 5378–5383.
(84) Smolin, N.; Winter, R. Molecular dynamics simulations of staphylococcal nuclease: Properties of water at the protein surface, *J. Phys. Chem. B* **2004**, *108*, 15928–15937.
(85) Kuffel, A.; Zielkiewicz, J. Why the solvation water around proteins is more dense than bulk water, *J. Phys. Chem. B* **2012**, *116*, 12113–12124.
(86) Conde, M. M.; Rovere, M.; Gallo, P. High precision determination of the melting points of water TIP4P/2005 and water TIP4P/Ice models by the direct coexistence technique, *J. Chem. Phys.* **2017**, *147*, 244506.
(87) Pettersson, L. G.; Nilsson, A. The structure of wa-

ter; from ambient to deeply supercooled, *J. Non. Cryst. Solids* **2015**, *407*, 399–417.
(88) Schlesinger, D.; Wikfeldt, K. T.; Skinner, L. B.; Benmore, C. J.; Nilsson, A.; Pettersson, L. G. The temperature dependence of intermediate range oxygen-oxygen correlations in liquid water, *J. Chem. Phys.* **2016**, *145*, 084503.
(89) Kauzmann, W. Some factors in the interpretation of protein denaturation, *Adv. Protein Chem.* **1959**, *14*, 1–63.
(90) Persson, F.; Söderhjelm, P.; Halle, B. The spatial range of protein hydration, *J. Chem. Phys.* **2018**, *148*, 215104.

TOC Graphic